

CANCER-LINKED GENES AS TARGETS FOR CHEMOTHERAPY

5

This application claims priority of U.S. Provisional Application 60/509,515, filed 8 October 2003, the disclosure of which is hereby
10 incorporated by reference in its entirety.

15

FIELD OF THE INVENTION

The present invention relates to methods of identifying cancer-target genes and screening such cancer-target genes and expression products for
20 involvement in the cancer initiation and facilitation process and the use of such genes for screening potential anti-cancer agents, including the design of small organic compounds and other molecules, and in the diagnosis of cancer.

25

BACKGROUND OF THE INVENTION

Cancer-linked genes are valuable in that they indicate genetic differences between cancer cells and normal cells, such as where a gene is
30 expressed in a cancer cell but not in a non-cancer cell, or where said gene is over-expressed or expressed at a higher level in a cancer as opposed to normal or non-cancer cell. In addition, the expression of such a gene in a normal cell but not in a cancer cell, especially of the same type of tissue, can indicate important functions in the cancerous process. For example, screening
35 assays for novel drugs are based on the response of model cell based

systems *in vitro* to treatment with specific compounds. Various measures of cellular response have been utilized, including the release of cytokines, alterations in cell surface markers, activation of specific enzymes, as well as alterations in ion flux and/or pH. Some such screens rely on specific genes, such as oncogenes (or gene mutations). In accordance with the present invention, cancer-target genes, and encoded polypeptides, have been identified. Such genes are useful in the diagnosing of cancer, the screening of anticancer agents and the treatment of cancer using such agents, especially in that these genes encode polypeptides that can act as markers, such as cell surface markers, thereby providing ready targets for anti-tumor agents such as antibodies, preferably antibodies complexed to cytotoxic agents, including apoptotic agents.

15

BRIEF SUMMARY OF THE INVENTION

In accordance with the present invention, there is provided herein a set of genes related to, or linked to, cancer, or otherwise involved in the cancer initiating and facilitating process and referred to as cancer-target genes, as well as polypeptides encoded by such genes.

In a particular embodiment, such genes are those corresponding to KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2 and which encode polypeptides.

More particularly, such genes whose expression is changed in cancerous, as compared to non-cancerous cells, from a specific tissue, for example, lung, where the gene would include a polynucleotide corresponding to one of the genes designated KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1,

ITPR2, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2* of genes substantially identical to said genes and/or encode the same or similar polypeptide or a polypeptide differing therefrom by conservative amino acid substitutions.

5 It is another object of the present invention to provide methods of using such characteristic genes as a basis for assaying the potential ability of selected chemical agents to modulate upward or downward the expression of said cancer characteristic, or related, or target genes.

10 It is a further object of the present invention to provide methods of detecting the expression, or non-expression, or amount of expression, of said characteristic gene, or portions thereof, as a means of determining the cancerous, or non-cancerous, status (or potential cancerous status) of selected cells as grown in culture or as maintained *in situ*.

15

 It is a still further object of the present invention to provide methods for treating cancerous conditions utilizing selected chemical agents as determined from their ability to modulate (i.e., increase or decrease) the characteristic gene, or its protein product.

20

 The present invention also relates to a method for treating cancer comprising contacting a cancerous cell with an agent having activity against an expression product encoded by one or more of the genes, which process may be conducted either *ex vivo* or *in vivo* and which product is disclosed
25 herein. Such agents may comprise an antibody or other molecule or portion that is specific for said expression product. In a preferred embodiment, the polypeptide product of such genes is a polypeptide encoded by one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, *ITPR2*, *KCNK1*, *KCNK5*,
30 *PRO2000*, *RFC2* and *RIPK2*.

DEFINITIONS

As used herein and except as noted otherwise, all terms are defined as given below.

5

The term "druggable" or "druggable domain" refers to a gene that encodes a protein domain known to be modulated by chemical compounds.

As used herein, the term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). It could also be produced recombinantly and subsequently purified. For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides, for example, those prepared recombinantly, could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. In one embodiment of the present invention, such isolated, or purified, polypeptide is useful in generating antibodies for practicing the invention, or where said antibody is attached to a cytotoxic or cytolytic agent, such as an apoptotic agent.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. As used herein, the terms "portion," "segment," and "fragment," when used in relation to polypeptides, refer to a continuous sequence of residues, such as amino acid residues, which sequence forms a subset of a larger sequence. For example, if a polypeptide were subjected to treatment with any of the common endopeptidases, such as trypsin or chymotrypsin, the oligopeptides resulting from such treatment would represent portions, segments or fragments of the starting polypeptide. When used in relation to a polynucleotides, such terms

refer to the products produced by treatment of said polynucleotides with any of the common endonucleases.

As used herein, the term "corresponding genes" refers to genes that
5 encode an RNA that is at least 90% identical, preferably at least 95% identical, most preferably at least 98% identical, and especially identical, to an RNA encoded by one of the genes of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*.

10

As used herein, the term "correspond" means that the gene has the same nucleotide sequence as a gene disclosed herein or that it encodes substantially the same RNA as would be encoded by the disclosed gene, the term "substantially" meaning at least 90% identical as defined elsewhere
15 herein and includes splice variants thereof.

The term "percent identity" or "percent identical," when referring to a sequence, means that a sequence is compared to a claimed or described sequence after alignment of the sequence to be compared (the "Compared
20 Sequence") with the described or claimed sequence (the "Reference Sequence"). The Percent Identity is then determined according to the following formula:

$$\text{Percent Identity} = 100 [1 - (C/R)]$$

25

wherein C is the number of differences between the Reference Sequence and the Compared Sequence over the length of alignment between the Reference Sequence and the Compared Sequence wherein (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or
30 amino acid in the Compared Sequence and (ii) each gap in the Reference Sequence and (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence,

constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid.

5

If an alignment exists between the Compared Sequence and the Reference Sequence for which the percent identity as calculated above is about equal to or greater than a specified minimum Percent Identity then the Compared Sequence has the specified minimum percent identity to the
10 Reference Sequence even though alignments may exist in which the hereinabove calculated Percent Identity is less than the specified Percent Identity.

The term "expression product" means that polypeptide or protein that is
15 the natural translation product of the gene and any nucleic acid sequence coding equivalents resulting from genetic code degeneracy and thus coding for the same amino acid(s).

The term "active fragment," when referring to a coding sequence, means
20 a portion comprising less than the complete coding region whose expression product retains essentially the same biological function or activity as the expression product of the complete coding region.

The term "primer" means a short nucleic acid sequence that is paired
25 with one strand of DNA and provides a free 3'-OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

The term "promoter" means a region of DNA involved in binding of RNA polymerase to initiate transcription. The term "enhancer" refers to a region of
30 DNA that, when present and active, has the effect of increasing expression of a different DNA sequence that is being expressed, thereby increasing the amount of expression product formed from said different DNA sequence.

The term "protein domain" refers to a discrete portion of a single polypeptide chain with its own function. The combination of domains in a single protein determines its overall function. Protein domains can act as independent units, to the extent that they can be excised from the chain, and still be shown to fold correctly, and often still exhibit biological activity. Another property of domains is that they are regions which are usually conserved during recombination events. This means that along a protein sequence, the domains will tend to be fairly well conserved, and conversely, the interdomain regions will be more divergent

As used herein, the term "conservative amino acid substitution" are defined herein as exchanges within one of the following five groups:

- I. Small aliphatic, nonpolar residues:
Ala, Gly;
- II. Negatively charged residues:
Asp, Glu
- III. Positively charged residues:
His, Arg, Lys
- IV. Large, aliphatic, nonpolar residues:
Met, Leu, Ile, Val, Cys
- V. Aromatic residues:
Phe, Tyr, Trp, Pro
- VI. Polar residues
Ser, Thr
- VII. Amides
Asn, Gln

SEQUENCE LISTING ON CD-ROM ONLY

The sequences disclosed herein as SEQ ID NO: 1-501 in the sequence listing are contained on compact disc (CD-ROM) only (denoted as Filename: 5 Avalon 213 (1,965 kB), 4 copies of which appear on discs denoted Copy 1, Copy 2, Copy 3 and CRF, and which discs were generated on 6 October 2004), which accompanies this application and the contents of said CD-ROMs are hereby incorporated by reference in their entirety. These sequence numbers also appear in Table 6 where all sequences are referred to as 10 consecutive Sequence ID Nos. for reference.

DETAILED SUMMARY OF THE INVENTION

15

The present invention relates to processes for identifying and/or utilizing cancer-target genes, and expression products of such genes, as targets for chemotherapeutic agents, especially anti-cancer agents.

20

Genes whose expression, or non-expression, or change in expression, are indicative of the cancerous or non-cancerous status of a given cell and whose expression is changed in cancerous, as compared to non-cancerous cells, from a specific tissue, are genes that are disclosed herein or that are identified by methods disclosed herein. These include genes having structural and/or functional similarity to the genes disclosed herein and include genes 25 that are substantially identical to said genes. In terms of nucleotide sequence, such genes are at least about 90% identical, preferably 95% identical, most preferably at least about 98% identical and especially where such gene is a gene of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, 30 CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2.

The genes disclosed herein according to the invention were identified within an amplified chromosomal region in a cancer cell line(s) and exhibit RNA over-expression in the cell line(s) and clinical tumor tissues by Affymetrix microarray analysis. Each disclosed gene contains sequences that encode a protein domain previously described as being modulated by chemical compounds.

Each such gene was identified in a cancer cell line(s) for which high-resolution comparative genomic hybridization (CGH) data and Affymetrix U133 chip expression data were generated. Each meets the following criteria:

- 1) At least 5 fold over-expressed in a cancer cell line(s) and mapped to a chromosomal region with a CGH ratio of 1.25 or above.
- 2) RNA expression level of at least 1.5 fold or higher in tumor tissue samples compared to corresponding normal tissue samples in a genetic database (with Gene Logic GX2000 database being a non-limiting example).
- 3) The gene encodes a protein domain known to be modulated by chemical compounds (i.e., a "druggable" domain). The genes identified herein represent a subset of all genes in these classes.

In accordance with the foregoing, the present invention relates to nucleotide sequences and derived polypeptides having the following characteristics:

Arbitrary Gene Name: KIAA1274
Description: KIAA protein (similar to mouse paladin)
UniGene: Hs.300646
Accession: AB033100
Gencarta No. AA553584 (Gene 2)
Cytogenetic location: 10q22.1
Affymetrix fragment: 231887_s_at
Affymetrix ID: 262356
Cell lines involved: SW620
Druggable domain: phosphatase

Arbitrary Gene Name: NEK6

Description: NIMA (never in mitosis gene a)-related kinase 6

UniGene: Hs.9625

5 Accession: BE616825

Gencarta No. T11445 (Gene 12)

Cytogenetic location: 9q33.3

Affymetrix fragment: 223158_s_at

Affymetrix ID: 253651

10 Cell lines involved: Colo205

Druggable domain: kinase

NIMA-related kinases (NEKs) are mammalian serine/threonine protein
kinases structurally related to Aspergillus NIMA (never in mitosis, gene A),
15 which play essential roles in mitotic signaling.

Arbitrary Gene Name: PAK2

Description: p21-activated kinase 2

20 UniGene: Hs.56974

Accession: BF796470

Gencarta No. Z26993 (Gene 17)

Cytogenetic location: 3q29

Affymetrix fragment: 208875_s_at

25 Affymetrix ID: 239505

Cell lines involved: HCC1954, HCC202, HCC70, MDA_MB453, T47D

Druggable domain: kinase

The p21-activated kinases (PAK) are critical effectors that link Rho GTPases
30 to cytoskeleton reorganization and nuclear signaling. The PAK proteins are a
family of serine/threonine kinases that serve as targets for the small GTP
binding proteins, CDC42 and RAC1, and have been implicated in a wide
range of biological activities.

35

Arbitrary Gene Name: PAK4

Description: p21-activated kinase 4

UniGene: Hs.20447

Accession: AF005046

40 Gencarta No. R09837 (Gene 8)

Cytogenetic location: 19q13.2

Affymetrix fragment: 33814_s_at

Affymetrix ID: 107335

Cell lines involved: HCC202, MDA_MB468

Druggable domain: kinase

The p21 activated kinases (PAK) are critical effectors that link Rho
5 GTPases to cytoskeleton reorganization and nuclear signaling. The PAK
proteins are a family of serine/threonine kinases that serve as targets for the
small GTP binding proteins, CDC42 and RAC1, and have been implicated in a
wide range of biological activities.

10

Arbitrary Gene Name: STK38L
Description: serine/threonine kinase 38 like
UniGene: Hs.184523
Accession: AW779556
15 Gencarta No. R14324 (Gene 9)
Cytogenetic location: 12p11.23
Affymetrix fragment: 212565_at
Affymetrix ID: 243092
Cell lines involved: HCC827, BEN
20 Druggable domain: kinase

25

Arbitrary Gene Name: ACP1
Description: acid phosphatase 1, soluble
UniGene: Hs.75393
Accession: BE872974
Gencarta No. HUMAAPA (Gene 6)
Cytogenetic location: 2p25.3
Affymetrix fragment: 201629_s_at
30 Affymetrix ID: 232293
Cell lines involved: BEN, NCI-H460, NCI-H522, MCF7, MDA-MB436, MDA-
MB468
Druggable domain: phosphatase

35

The product of this gene belongs to the phosphotyrosine protein
phosphatase family of proteins. It functions as an acid phosphatase and a
protein tyrosine phosphatase by hydrolyzing protein tyrosine phosphate to
protein tyrosine and orthophosphate. This gene is genetically polymorphic,
40 and three common alleles segregating at the corresponding locus give rise to
six phenotypes. Each allele appears to encode at least two electrophoretically

different isozymes, Bf and Bs, which are produced in allele-specific ratios. Three transcript variants encoding distinct isoforms have been identified for this gene (Bryson et al., Genomics 1995 Nov 20;30(2):133-40).

5

Arbitrary Gene Name: *ARHC*

Description: ras homolog gene family, member C (hypothetical protein MGC19531)

10 UniGene: Hs.446391

Accession: AW117553

GenCarta No. AA383349 (Gene 1)

Cytogenetic location: 1p13.2

Affymetrix fragment: 229484_at

15 Affymetrix ID: 259953

Cell lines involved: HCC202, MDA_MB468

Druggable domain: phosphatase

20 *ARHC* (UniGene Hs.179735) sits next to the gene for hypothetical protein MGC19531. The two sequences are in close proximity and they are annotated as the same gene in GenCarta, but they are listed as two distinct genes in UCSC Goldenpath. The Affymetrix fragment maps within the sequence for hypothetical protein MGC19531, but we are inferring that this
25 fragment is detecting expression for *ARHC*.

ARHC encodes a ras-related GTP binding protein of the rho subfamily, member C (RhoC) that regulates remodeling of the actin cytoskeleton during cell morphogenesis and motility. Up regulation of RhoC through increased
30 expression of *ARHC* has been reported in breast, ovarian and pancreatic cancer as well as melanoma and has been associated with progression to a metastatic phenotype in each cancer type (van Golen et al., Cancer Res. 2000 Oct 15;60(20):5832-8, Horiuchi A et al. Lab Invest. 2003 83(6):861-70, Suwa et al. Br J Cancer. 1998 77(1):147-52, Clark et al., Nature, 2000
35 406(6795):532-5).

Arbitrary Gene Name: *CDC6*

Description: CDC6 cell division cycle 6 homolog (*S. cerevisiae*)

UniGene: Hs.69563

Accession: U77949

Gencarta No. T83032 (Gene 16)

5 Cytogenetic location: 17q21.3

Affymetrix fragment: 203967_at

Affymetrix ID: 234629

Cell lines involved: NCI-H522, NCI-H23

Druggable domain: AAA ATPase

10

Yan et al. [Proc Nat Acad Sci 94:142-147 (1998)] showed that *CDC6* is expressed selectively in proliferating but not quiescent mammalian cells, both in culture and within tissues in intact animals. During the transition from a growth-arrested to a proliferative state, transcription of mammalian *CDC6* is regulated by E2F proteins as revealed by a functional analysis of the promoter and by the ability of exogenously expressed E2F proteins to stimulate endogenous *CDC6*. Immunodepletion of *CDC6* protein by microinjection of anti-*CDC6* antibody blocked initiation of DNA replication in a human tumor cell line. The authors concluded that expression of human *CDC6* is regulated in response to mitogenic signals through transcriptional control mechanisms involving E2F proteins, and that *CDC6* protein is required for initiation of DNA replication in mammalian cells.

25

Arbitrary Gene Name: *CDK7*

Description: Cyclin-dependent kinase 7 (MO15 homolog, *Xenopus laevis*)

UniGene: Hs.184298

30 Accession: X77743

Gencarta No. F02366 (Gene 4)

Cytogenetic location: 5q13.2

Affymetrix fragment: 211297_s_at

Affymetrix ID: 241855

35 Cell lines involved: SW620

Druggable protein domain: kinase

The protein encoded by *CDK7* is a member of the cyclin-dependent protein kinase (CDK) family, which are known to be important regulators of cell cycle progression. This protein forms a trimeric complex with cyclin H and

40

MAT1, which functions as a Cdk-activating kinase (CAK) (Fisher and Morgan, Cell 78:713-724, 1994). It is an essential component of the transcription factor IIH (TFIIH) that is involved in transcription initiation and DNA repair (Shiekhataar et al., Nature 374: 283-287, 1995). This protein is thought to
5 serve as a direct link between the regulation of transcription and the cell cycle.

Arbitrary Gene Name: CDKN3
10 Description: cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)
UniGene: Hs.84113
Accession: AF213033
Gencarta No. HUMPTPB (Gene 7)
15 Cytogenetic location: 14q22.2
Affymetrix fragment: 209714_s_at
Affymetrix ID: 240337
Cell lines involved: NCI-H460, HCC827, NCI-H23, HCC202, MCF7, MDA-MB453, T47D
20 Druggable domain: phosphatase

The protein encoded by this gene is a human dual specificity protein phosphatase that was identified as a cyclin-dependent kinase inhibitor, and
25 has been shown to interact with and dephosphorylate CDK2 kinase and thus prevent the activation of CDK2 kinase. The gene has been reported to be deleted, mutated, or overexpressed in several kinds of cancers.

Lee et al. [Mol Cell Biol. 2000 Mar;20(5):1723-32] identified the *CDKN3*
30 as an overexpressed gene in breast and prostate cancer by using a phosphatase domain-specific differential-display PCR strategy. They report in normal cells, CDKN3 protein is primarily found in the perinuclear region, but in tumor cells, a significant portion of the protein is found in the cytoplasm. Blocking *CDKN3* expression by antisense in a tetracycline-regulatable system
35 resulted in a reduced population of S-phase cells and reduced Cdk2 kinase activity. Furthermore, lowering *CDKN3* expression led to inhibition of the transformed phenotype, with reduced anchorage-independent growth and tumorigenic potential in athymic nude mice. They suggest that therapeutic

intervention might be aimed at repression of *CDKN3* gene overexpression in human breast and prostate cancer.

Yeh et al. [Cancer Res. 2000 Sep 1;60(17):4697-4700] analyzed
5 *CDKN3* mRNA in hepatocellular carcinoma by reverse transcription-PCR (RT-PCR), followed by cloning and sequencing. They found aberrant *CDKN3* transcripts in hepatocellular tumors and showed mutant proteins were defective in interacting with Cdk2.

10

Arbitrary Gene Name: *CRK7*

Description: CDC2-related protein kinase 7

UniGene: Hs.278346

Accession: AI651265

15 Gencarta No. T60764 (Gene 14)

Cytogenetic location: 17q12

Affymetrix fragment: 225697_at

Affymetrix ID: 256169

Cell lines involved: HCC1954, HCC202, SKBR3

20 Druggable domain: kinase

Ko et al. [Journal of Cell Science, 114,2591-2603 (2001)] isolated and characterized CrkRS, CDC2-related kinase 7, as a novel human protein with
25 an arginine/serine-rich (RS) domain that is most closely related to the cyclin-dependent kinase family. They report CrkRS is a 1490 amino acid protein where the protein kinase domain is 89% identical to CHED protein kinase. CrkRS has extensive proline-rich regions that match the consensus for SH3 and WW domain binding sites and RS domain that is predominantly found in
30 splicing factors. The authors describe CrkRS as a novel, conserved link between the transcription and splicing machinery of a cell.

Arbitrary Gene Name: *DUSP16*
35 Description: dual specificity phosphatase 16
UniGene: Hs.20281
Accession: AB051487
Gencarta No. T23935 (Gene 13)
Cytogenetic location: 12p13.2
40 Affymetrix fragment: 224832_at

Affymetrix ID: 255305
Cell lines involved: HCC827
Druggable domain: phosphatase

5

Mitogen-activated protein kinase (MAPK) phosphatases (MKPs) negatively regulate MAPK activity. DUSP16 is a dual specificity phosphatase that functions as a MAPK phosphatase, also known as MKP7. Masuda et al. [J Biol Chem. 2001 276(42):39002-11] showed that MAPK7 behaves as a nuclear shuttle for c-Jun terminal kinase (JNK) group of MAPKs as well as a phosphatase.

10

Arbitrary Gene Name: *FIGNL1*

15

Description: fidgetin-like 1

UniGene: Hs.137516

Accession: AA805691

Gencarta No. H61320 (Gene 5)

Cytogenetic location: 7p12.2

20

Affymetrix fragment: 222843_at

Affymetrix ID: 253337

Cell lines involved: SW620, BEN, HCC827

Druggable domain: AAA ATPase

25

Arbitrary Gene Name: GUK1

Description: guanylate kinase 1

30

UniGene: Hs.3764

Accession: BC006249

Gencarta No. T08090 (Gene 11)

Cytogenetic location: 1q42.13

Affymetrix fragment: 200075_s_at

35

Affymetrix ID: 231232

Cell lines involved: HCC202, MDA_MB436, MDA_MB453, MDA_MB468

Druggable domain: kinase

40

Guanylate kinase catalyzes the phosphorylation of either GMP to GDP or dGMP to dGDP and is an essential enzyme in nucleotide metabolism pathways. There are several isoforms, GUK2 and GUK3, determined by different loci. Brady et al. [J Biol Chem. 1996 271(28):16734-40] stated that

the guanylate kinases are targets for cancer chemotherapy and are inhibited by the drug 6-thioguanine. They report a model of the tertiary structure designed to be used in the development of chemotherapy drugs.

5

Arbitrary Gene Name: *ITPR2*

Description: inositol 1,4,5-triphosphate receptor, type 2

UniGene

Accession: D26350

10

Gencarta No. Z38709 (Gene 18)

Cytogenetic location: 12p11.23

Affymetrix fragment: 211360_s_at

Affymetrix ID: 241911

Cell lines involved: BEN, HCC827

15

Druggable domain: Ion transport

Arbitrary Gene Name: *KCNK1*

20

Description: potassium channel, subfamily K, member 1

UniGene: Hs.79351

Accession: U90065

Gencarta No. Z39663 (Gene 19)

Cytogenetic location: 1q42.2

25

Affymetrix fragment: 204678_s_at

Affymetrix ID: 235340

Cell lines involved: HCC202, HCC70, MDA_MB436, MDA_MB453, MDA_MB468

Druggable domain: potassium channel

30

This gene encodes one of the members of the superfamily of potassium channel proteins containing two pore-forming P domains and 4 transmembrane segments. Potassium channels are functionally important to a large number of cellular processes including maintenance of the action potential, muscle contraction, hormone secretion, osmotic regulation and ion flow.

35

40

Arbitrary Gene Name: *KCNK5*

Description: potassium channel, subfamily K, member 5

UniGene: Hs.127007

Accession: AI678413
 Gencarta No. R25184 (Gene 10)
 Cytogenetic location: 6p21.2
 Affymetrix fragment: 69854_at
 5 Affymetrix ID: 153971
 Cell lines involved: Colo201, Colo205
 Druggable domain: K⁺ channel

10 This gene encodes one of the members of the superfamily of potassium channel proteins containing two pore-forming P domains.

15 Arbitrary Gene Name: *PRO2000*
 Description: Hypothetical protein MGC5254
 UniGene: Hs.222088
 Accession: AI925583
 Gencarta No. Z44462 (Gene 20)
 20 Cytogenetic location: 8q24.13
 Affymetrix fragment: 222740_at
 Affymetrix ID: 253234
 Cell lines involved: BT549, HCC1954, HCC202, HCC70, Hs578t, MCF7, MDA_MB231, MDA_MB436, MDA_MB453, SKBR3, T47D, Colo201,
 25 HCT116, SW620, HT29, HCC827, NCI-H23, NCI-H460
 Druggable domain: AAA ATPase

A large family of ATPases has been described, whose key feature is
 30 that they share a conserved region of about 220 amino acids that contains an ATP-binding site. The protein encoded by *PRO2000* contains two AAA (ATPases Associated with diverse cellular Activities) domains as well as a bromodomain. AAA family proteins often perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes.
 35 The exact function of the *PRO2000* protein is unknown.

Fellenberg et al. [Int J Cancer 105(5):636-643 (2003)] report *PRO2000* is up-regulated >2 fold in osteosarcoma cell line (Saos-2) following treatment with cisplatin, methotrexate and doxorubicin.

40

Arbitrary Gene Name: RFC2
 Description: replication factor C (activator 1) 2, 40kDa
 UniGene: Hs.139226
 Accession: M87338

- 5 Gencarta No. T62520 (Gene 15)
 Cytogenetic location: 7q11.23
 Affymetrix fragment: 1053_at
 Affymetrix ID: 113880
 Cell lines involved: HCC827, NCI_H23, NCI_H522
 10 Druggable domain: AAA ATPase

The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and replication factor C (RFC). RFC, also called
 15 activator 1, is a protein complex consisting of five distinct subunits of 145, 40, 38, 37, and 36.5 kD. This gene encodes the 40 kD subunit, which has been shown to be responsible for binding ATP. Alternatively spliced transcript variants encoding distinct isoforms have been described.

20

Arbitrary Gene Name: RIPK2
 Description: receptor-interacting serine-threonine kinase 2
 25 UniGene: Hs.103755
 Accession: AF064824
 Gencarta No. D61791 (Gene 3)
 Cytogenetic location: 8q21.3
 Affymetrix fragment: 209545_s_at
 30 Affymetrix ID: 240173
 Cell lines involved: HCT116
 Druggable domain: kinase

35 The methods of the invention utilize these genes, designated as KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2.

40 The genes disclosed herein may be used in any of the methods of the invention and modulation, as used herein, may include modulation of

the gene, such as an increase or decrease in transcription or translation, and may include differences in the amount and/or the rate of production of RNA and/or polypeptide. Such modulation may affect any of the transcripts disclosed for the genes of the invention or any of the encoded polypeptides, as identified in Table 6. Antibodies useful in the invention would include those specific for any of the polypeptides encoded by these genes, especially any polypeptide whose sequences are provided in Figure 1, as identified in Table 6. A brief summary of these genes identified by their respective GenBank Accession Nos. is provided in Table 1.

10

Table 1. Brief summary of cancer target genes.

Accession	unigene	affy	Description
AI651265	Hs.278346	256169	CDC2-related protein kinase 7
U77949	Hs.69563	234629	CDC6 cell division cycle 6 homolog (S. cerevisiae)
X77743	Hs.184298	241855	cyclin-dependent kinase 7 (MO15 homolog, Xenopus laevis, cdk-activating kinase)
AA805691	Hs.137516	253337	fidgetin-like 1
AI925583	Hs.222088	253234	hypothetical protein MGC5254
D26350		241911	inositol 1,4,5-triphosphate receptor, type 2
	Hs.406293	253470	neurotrophic tyrosine kinase, receptor, type 1
BE616825	Hs.9625	253651	NIMA (never in mitosis gene a)-related kinase 6
BF796470	Hs.56974	239505	p21 (CDKN1A)-activated kinase 2
AF005046	Hs.20447	107335	p21(CDKN1A)-activated kinase 4
U90065	Hs.79351	235340	potassium channel, subfamily K, member 1
AF064824	Hs.103755	240173	receptor-interacting serine-threonine kinase 2
M87338	Hs.139226	113180	replication factor C (activator 1) 2, 40kDa
AW779556	Hs.184523	243092	serine/threonine kinase 38 like
AI678413	Hs.127007	153971	potassium channel, subfamily K, member 5
AF213033	Hs.84113	240337	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)
BE872974	Hs.75393	232293	acid phosphatase 1, soluble
AW117553	Hs.446391	259953	hypothetical protein MGC19531 (ras homolog gene family, member C)
BC006249	Hs.3764	231232	guanylate kinase 1
AB033100	Hs.300646	262356	KIAA protein (similar to mouse paladin)
AB051487	Hs.20281	255305	dual specificity phosphatase 16

Table 2 describes the location of the cancer target genes of the present invention while Table 3 describes primers used to locate these genes. An additional set of primers is provided in Table 5 while additional gene data is provided in Table 4.

5

The nucleotides and polypeptides, as gene products, used in the methods of the present invention may comprise a recombinant polynucleotide or polypeptide, a natural polynucleotide or polypeptide, or a synthetic polynucleotide or polypeptide, preferably a recombinant polynucleotide or polypeptide.

10

Table 2. Chromosome Location of Cancer Target Genes

Accession	Chromosome	band	Description	indication	Primer
AI651265	chr17	q12	kinase		PR3869
U77949	chr17	q21.2	AAA ATPase	breast	PR3870
X77743	chr5	q13.2	kinase	ovary	PR3871
AA805691	chr7	p12.2	AAA ATPase	lung	PR3872
AI925583	chr8	q24.13	AAA ATPase		PR3873
D26350	chr12	p11.23	ion transport		PR3874
	chr1	q21.3	tk	melanoma and LN mets	PR3875
BE616825	chr9	q33.3	kinase	sarcoma	PR3876
BF796470	chr3	q29	kinase	ovary	PR3877
AF005046	chr19	q13.2	kinase		PR3878
U90065	chr1	q42.2	K ⁺ channel	pancreas	PR3879
AF064824	chr8	q21.3	kinase	ovary	PR3880
M87338	chr7	q11.23	AAA ATPase		PR3881
AW779556	chr12	p11.23	kinase	pancreas	PR3882
AI678413	chr6	p21.2	K ⁺ channel		PR3883
AF213033	chr14	q22.2			PR3884
BE872974	chr2	p25.3			PR3885
AW117553	chr1	p13.2	Phosphatase	breast	PR3886
BC006249	chr1	q42.13			PR3887
AB033100	chr10	q22.1			PR3888
AB051487	chr12	p13.2			PR3889

15

Fragments of such polynucleotides and polypeptides as are disclosed herein may also be useful in practicing the processes of the present invention. For example, a fragment, derivative or analog of a polypeptide encoded by one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*,
5 *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2* may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in
10 which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence
15 which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The genes and gene products useful in practicing the methods of the
20 present invention may likewise be obtained in an isolated or purified form. In addition, the polypeptide disclosed herein as being useful in practicing the processes of the invention include different types of proteins in terms of function so that, as recited elsewhere herein, some are enzymes, some are transcription factors and other may be cell surface receptors. Precisely how such cancer-
25 linked proteins are used in the processes of the invention may thus differ depending on the function and cellular location of the protein and therefore modification, or optimization, of the methods disclosed herein may be desirable in light of said differences. For example, a cell-surface receptor is an excellent target for cytotoxic antibodies whereas a transcription factor or enzyme is a
30 useful target for a small organic compound with anti-neoplastic activity.

Table 3. Primers used to Identify Genes

Accession	Left primer 1	Right primer 1
AI651265	TCTTGGGCATCTCAACG	CACATAAAGCAGGTTGTAGAACG
U77949	TATTCAGCTGGCATTAGAGAGC	ACACTTGCAAGCACATTGGC
X77743	TGGACAACATTTTACTACTGAGGG	TAAGTTTTCCCTAATGCATTTTCA
AA805691	TTAGTCCTGATAAAAATTAAAAACC	AGAAAGTGCCTTCCTGATG
AI925583	TAAACCATTGTTGGGATGGCAT	TGGAACAAGCTGTTAACACCC
D26350	CTCTGAGGACATTCCCGTTAGAA	CAGGTGTTTCAAGGAAGAGGAAA
	ATGACATGGGGCTTGC	AACACCTGAGGGGGCTT
BE616825	TCTTCATGAATTCTAAGTAACTC	GTCACCTGACTTCATGACA
BF796470	TAACAAGCGATTCTAAACCACC	ATGGATGCAAATTCTTTAAGCA
AF005046	AACTAACTCGAGGCAGGGGT	CTGCCCTTATTGGGGGAC
U90065	TTCCCCTTATTTTATTGTAGCAA	GGTTTATGTGTACTGGTTTGCA
AF064824	AAATGGGGACAGGAAGCC	GCTTAATTGCCCTACAAAGGG
M87338	CAACAAACACTGCAAGGCTT	TCTCCATCCTGGGGAAAAA
AW779556	TGCCACCAAACATTTTTGA	ATGTGAGGGGATATTGCTGC
AI678413	AACTACTACACACAGAAGCTGC	AAGCCAGCTTCAGATGTATAT
AF213033	CCATGTCTGAAATGTCAGTTCTC	AAAACCTTAGGAATATCTGCACATG
BE872974	TCAGAGGCAAAGTGGTTCAG	AATCAGTCGTTGGCACCTTC
AW117553	TCTTGACACATACGAAGCC	GTAGAAGCAGAGTCCCTGG
BC006249	AGGCTTGCTGTCTGTTCTCG	TTTATTAGGATGTCAGCCCTGG
AB033100	CTTCTCCTCAGTCTCAAACCCAA	ATCCATCTCTCTGACAGTGCTGA
AB051487	ATCCCATTTTAAACAATTCTTTGA	GCTGAACCACCAGGAACCT

Table 4. Further Description of Cancer Target Genes

Accession			UniSTS	Gencarta Name
AI651265	PR3890	CRK7	SHGC-58832	T60764
U77949	PR3891	CDC6	RH70424	T83032
X77743	PR3892	CDK7	SHGC-149358	F02366
AA805691	PR3893	FIGNL1	RH103568	H61320
AI925583	PR3894	PRO2000	RH80934	Z44462
D26350	PR3895	ITPR2	SHGC-106565	Z38709
	PR3896		SHGC-69193	
BE616825	PR3897	NEK6	RH62928	T11445
BF796470	PR3898	PAK2	SHGC-35416	Z26993
AF005046	PR3899	PAK4	RH39107	R09837
U90065	PR3900	KCNK1	55164	Z39663
AF064824		RIPK2		D61791
M87338	PR3901	RFC2	47404	T62520
AW779556	PR3902	STK38L	182659	R14324
AI678413	PR3903	KCNK5	83108	R25184
AF213033	PR3904	CDKN3	24341	HUMPTPB
BE872974	PR3905	ACP1	91295	HUMAAPA
AW117553	PR3906	ARHC	RH49960	AA383349
BC006249	PR3907	GUK1	38548	T08090
AB033100	PR3908	KIAA1274	148013	AA553584
AB051487	PR3909	DUSP16	85676	T23935

Table 5. Additional Primers for Cancer Target Genes

Accession	Left primer 2	Right primer 2
AI651265	GTGGGGCCCAATAACTCAA	TTTTGAATCTGGCCTTGCCT
U77949	TTATGACCCCAACGCC	AAGCAAGTCCACATGGAG
X77743	CAGAGGTTCCCTCTTAAAAATTCA	AAAGTGAAGTATTGGCTGGGC
AA805691	CCATCCATGGAATCCTAGACA	TTATCCTACCACTTTGCGGG
AI925583	AAGAGTTGGCCAACTTCAACTATT	TGTCATGTCCGCCTAATTGA
D26350	CAAAGCCTCAAGACCTTTTTCAA	AAGGTACCAGCTAAACCTCTTTGC
	GAGAAAGGGAGGGATCGTTC	TGTGAGGGGCTATGCTGG
BE616825	TTCCACTTTATCCCTTTACAACA	GGCTTATGCTAACAGGAGACTTG
BF796470	TCACTGCTGTGGCCTCATAC	TCAGTCCACAATTCCTTCTGG
AF005046	GGGGGACGCTGTCATTAC	TTCCCAGTACCGCAGAGCC
U90065	GGTCCTCTACTTCCACAT	GCTCTCTGAATTTTGTATT
AF064824		
M87338	GCAGAGACTTCACTGACTGAC	TGACCTCAGGTGATCCACCTG
AW779556	TTTAGCAAACTTGGAGCTGGAG	AAAACCATTCTCTACTAACTACCCCC
AI678413	TTTTGCAAGGCAACTGAGG	GATACGGCAGCCTCTACTGC
AF213033	CACATGGCCTAGTAGTTTGG	GTTCCAACTGCTTAGATCAGC
BE872974	TGAACAAAGAGCTGGGCTTT	ACTGAGGCAGGTTCTGTGC
AW117553	AGCCTGTAGCCTTTATCCATG	CTTCTGGCTCACAGGAAAATG
BC006249	CTGCTCTTACCTGGGGTTG	GAGCCACAGAGGAGTGAAGG
AB033100	ACATGTGCCCTACACACAC	AGCTGTCACATAAATAGAACCC
AB051487	ATCAGACATTCTCAAGTTTCACACA	GGACCATGGCCAAGAGAAG

- 5 Expression products of the genes disclosed herein for use in the methods of the invention may be in an isolated form.

Methods of producing vectors comprising genes disclosed herein, or recombinant cells expressing such genes, are well known to those skilled in the molecular biology art. See, for example, Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, N.Y., (1989), Wu et al, *Methods in Gene Biotechnology* (CRC Press, New York, NY, 1997), and *Recombinant Gene Expression Protocols*, in *Methods in Molecular Biology*, Vol. 62, (Tuan, ed., Humana Press, Totowa, NJ, 1997), the disclosures of which are hereby incorporated by reference.

In one aspect, the present invention relates to a method for identifying a cancer-target gene, comprising:

a) identifying a gene that is at least 5 fold over-expressed in a cancer cell line and that maps to a chromosomal region with a CGH ratio of at least 1.25;

b) determining an RNA expression level of said gene of at least 1.5 fold in a tumor tissue compared to corresponding normal tissue in a genetic database,

c. determining that said gene encodes a protein domain known to be modulated, or shown to be modulated, by chemical compounds

wherein a gene that meets the criteria of a, b and c is considered to be a cancer-target gene,

thereby identifying a cancer-target gene.

The present invention also relates to a set of cancer-target genes identified using such methods. The genes disclosed herein form such a set. In addition, subsets of such sets are specifically contemplated by the invention.

In another aspect, the present invention relates to a method for identifying an agent that modulates the activity of a cancer-target gene comprising:

(a) contacting a test compound with a cell containing a polynucleotide that corresponds to a gene that has the properties of a, b and c of claim 1 and under conditions promoting the expression of said gene, and

(b) determining a difference in expression of said gene relative to when said test compound is not present wherein said difference indicates gene modulating activity,

thereby identifying an agent that modulates the activity of a cancer-related gene.

In a preferred embodiment, said gene was first identified as a cancer target gene using one or more of the methods of the invention.

In another preferred embodiment, the gene is a gene selected from the group consisting of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, CDKN3, *CRK7*, DUSP16, *FIGNL1*, GUK1, *ITPR2*, KCNK1, 5 KCNK5, *PRO2000*, RFC2 and RIPK2. These 20 genes are contained in Table 6 where they are described in terms of a consensus sequence along with identified polynucleotide transcripts and polypeptides.

In the methods of the invention, expression may be determined by 10 determining transcription (to form RNA), as by measuring the rate or amount of RNA formed, or translation (to form protein), such as where antibodies may be used to determine the amount of polypeptide or protein formed from the gene in question or where the activity of such protein is determined, such as where the protein is an enzyme and the amount of enzyme activity can be 15 determined.

In one preferred embodiment, the cell is a cancer cell and the determined difference in expression is a decrease in expression. In another embodiment, the cell is a recombinant cell, such as one comprising a gene as 20 disclosed herein, and the difference in expression is a decrease in expression.

The present invention also relates to a method for identifying an anti-neoplastic agent comprising contacting a cell exhibiting neoplastic activity with 25 a compound first identified as a cancer target gene modulator using one of the methods of the invention and detecting a decrease in said neoplastic activity after said contacting compared to when said contacting does not occur. In preferred embodiments, the neoplastic activity is accelerated cellular replication. In another preferred embodiment, the decrease in neoplastic 30 activity results from the death of the cell. In a preferred embodiment, the compound is one that modulates, preferably inhibits, a gene disclosed herein, most preferably a gene identified in Table 6.

The present invention also relates to a method for identifying an anti-neoplastic agent comprising administering to an animal exhibiting a cancerous condition an effective amount of a cancer target gene modulating agent by a method of the invention and detecting a decrease in said cancerous condition. In a preferred embodiment, the compound is one that modulates, preferably inhibits, a gene disclosed herein, most preferably a gene identified in Table 6.

In accordance with the present invention, model cellular systems using cell lines, primary cells, or tissue samples are maintained in growth medium and may be treated with compounds that may be at a single concentration or at a range of concentrations. At specific times after treatment, cellular RNAs are isolated from the treated cells, primary cells or tumors, which RNAs are indicative of expression of selected genes. The cellular RNA is then divided and subjected to analysis that detects the presence and/or quantity of specific RNA transcripts, which transcripts may then be amplified for detection purposes using standard methodologies, such as, for example, reverse transcriptase polymerase chain reaction (RT-PCR), etc. The presence or absence, or levels, of specific RNA transcripts are determined from these measurements and a metric derived for the type and degree of response of the sample to the treated compound compared to control samples.

In accordance with the foregoing, there are thus disclosed herein methods for using a cancer-linked or cancer-target gene sequence (such as that of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*) whose expression is, or can be, as a result of the methods of the present invention, linked to, or used to characterize, the cancerous, or non-cancerous, status of the cells, or tissues, to be tested. Thus, the processes of the present invention identify novel anti-neoplastic agents based on their alteration of expression of the polynucleotide sequence

disclosed herein in specific model systems. The methods of the invention may therefore be used with a variety of cell lines or with primary samples from tumors maintained *in vitro* under suitable culture conditions for varying periods of time, or *in situ* in suitable animal models.

5

More particularly, genes have been identified that are expressed at a level in cancer cells that is different from the expression level in non-cancer cells. In one instance, the identified genes are expressed at higher levels in cancer cells than in normal cells.

10

The genes useful in the methods of the invention can include fully operational genes with attendant control or regulatory sequences or merely a polynucleotide sequence encoding the corresponding polypeptide or an active fragment or analog thereof.

15

In one embodiment of the present invention, said gene modulation is downward modulation, so that, as a result of exposure to the chemical agent to be tested, one or more genes of the cancerous cell will be expressed at a lower level (or not expressed at all) when exposed to the agent as compared to the expression when not exposed to the agent.

20

In a preferred embodiment a selected set of said genes are expressed in the reference cell, including the gene(s) identified for use according to the present invention, but are not expressed in the cell to be tested as a result of the exposure of the cell to be tested to the chemical agent. Thus, where said chemical agent causes the gene, or genes, of the tested cell to be expressed at a lower level than the same genes of the reference, this is indicative of downward modulation and indicates that the chemical agent to be tested has anti-neoplastic activity.

25

30

The genes identified by the present disclosure are considered "cancer-related" genes, or cancer-target" genes, as this term is used herein, and

include genes expressed at higher levels (due, for example, to elevated rates of expression, elevated extent of expression or increased copy number) in cancer cells relative to expression of these genes in normal (i.e., non-cancerous) cells where said cancerous state or status of test cells or tissues has been determined by methods known in the art, such as by reverse transcriptase polymerase chain reaction (RT-PCR) as described in the Example below. In specific embodiments, this relates to the genes of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2.

The genes disclosed herein may be genomic in nature and thus represent an actual gene as found in nature, such as a human gene, or may be a cDNA sequence derived from a messenger RNA (mRNA) and thus represent contiguous exonic sequences derived from a corresponding genomic sequence or they may be wholly synthetic in origin for purposes of practicing the processes of the invention. Because of the processing that may take place in transforming the initial RNA transcript into the final mRNA, the genes disclosed herein may represent less than the full genomic nucleotide sequence. They may also represent sequences derived from ribosomal and transfer RNAs. Consequently, the genes present in the cell (and representing the genomic sequences) and the sequences of genes disclosed herein, which are mostly cDNA sequences, may be identical or may be such that the cDNAs contain less than the full genomic sequence. Such genes and cDNA sequences are still considered as corresponding to genes disclosed herein because they both encode similar RNA sequences. Thus, by way of non-limiting example only, a gene that encodes an RNA transcript, which is then processed into a shorter mRNA, is deemed to encode both such RNAs and therefore encodes an RNA complementary to (using the usual Watson-Crick complementarity rules), or that would otherwise be encoded by, a cDNA (for example, a sequence as disclosed herein). Thus, the sequences of genes disclosed herein correspond to genes contained in the cancerous or normal cells used to determine relative levels of expression because they represent

the same sequences or are complementary to RNAs encoded by these genes. Such genes also include different alleles and splice variants that may occur in the cells used in the processes of the invention.

5 The genes of the invention "correspond to" the genes of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, DUSP16, *FIGNL1*, GUK1, *ITPR2*, KCNK1, KCNK5, *PRO2000*, RFC2 and RIPK2 if the gene encodes an RNA (processed or unprocessed, including naturally occurring splice variants and alleles) that is at least 90% identical, preferably at least 95% identical, most preferably at least 98% identical to, and especially identical to, an RNA that would be encoded by, or be complementary to, such as by hybridization with, a polynucleotide having the indicated sequence. In addition, genes including sequences at least 90% identical to a genes of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, 10 *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, DUSP16, *FIGNL1*, GUK1, *ITPR2*, KCNK1, KCNK5, *PRO2000*, RFC2 or RIPK2, preferably at least about 95% identical to such a sequence, more preferably at least about 98% identical to such sequence and most preferably comprising such sequence are specifically contemplated by all of the processes of the present invention as being genes that correspond to these sequences. In addition, genes encoding the same proteins as any of these genes, regardless of the percent identity of such sequences, are also specifically contemplated by any of the methods of the present invention that rely on any or all of said sequences, regardless of how they are otherwise described or limited. Thus, any such sequences are 20 available for use in carrying out any of the methods disclosed according to the invention.

Such genes will also encode the same or similar polypeptide sequence as the genes KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, 30 *CDK7*, *CDKN3*, *CRK7*, DUSP16, *FIGNL1*, GUK1, *ITPR2*, KCNK1, KCNK5, *PRO2000*, RFC2 and RIPK2 but may include differences in such amino acid sequences where such differences are limited to conservative amino acid

substitutions, such as where the same overall three dimensional structure, and thus the same antigenic character, is maintained. Thus, amino acid sequences may be within the scope of the present invention where they react with the same antibodies that react with polypeptides encoded by genes disclosed herein, preferably KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2.

The present invention also relates to methods of assaying potential antitumor agents based on their modulation of the expression of the disclosed genes according to the invention and methods for diagnosing cancerous, or potentially cancerous, conditions as a result of the patterns of expression of the a gene disclosed herein as well as related genes based on common expression or regulation of such genes.

15

In carrying out the foregoing assays, relative antineoplastic activity may be ascertained by the extent to which a given chemical agent modulates the expression of genes present in a cancerous cell. Thus, a first chemical agent that modulates the expression of a gene associated with the cancerous state (i.e., a gene that includes one of the genes of the invention as disclosed herein and present in cancerous cells) to a larger degree than a second chemical agent tested by the assays of the invention is thereby deemed to have higher, or more desirable, or more advantageous, anti-neoplastic activity than said second chemical agent.

25

The gene expression to be measured is commonly assayed using RNA expression as an indicator. Thus, the greater the level of RNA (messenger RNA) detected the higher the level of expression of the corresponding gene. Thus, gene expression, either absolute or relative, is determined by the relative expression of the RNAs encoded by such genes.

30

RNA may be isolated from samples in a variety of ways, including lysis and denaturation with a phenolic solution containing a chaotropic agent (e.g., triazol) followed by isopropanol precipitation, ethanol wash, and resuspension in aqueous solution; or lysis and denaturation followed by isolation on solid support, such as a Qiagen resin and reconstitution in aqueous solution; or lysis and denaturation in non-phenolic, aqueous solutions followed by enzymatic conversion of RNA to DNA template copies.

Normally, prior to applying the processes of the invention, steady state RNA expression levels for the genes, and sets of genes, disclosed herein will have been obtained. It is the steady state level of such expression that is affected by potential anti-neoplastic agents as determined herein. Such steady state levels of expression are easily determined by any methods that are sensitive, specific and accurate. Such methods include, but are in no way limited to, real time quantitative polymerase chain reaction (PCR), for example, using a Perkin-Elmer 7700 sequence detection system with gene specific primer probe combinations as designed using any of several commercially available software packages, such as Primer Express software., solid support based hybridization array technology using appropriate internal controls for quantitation, including filter, bead, or microchip based arrays, solid support based hybridization arrays using, for example, chemiluminescent, fluorescent, or electrochemical reaction based detection systems.

The gene patterns indicative of a cancerous state need not be characteristic of every cell found to be cancerous. Thus, the methods disclosed herein are useful for detecting the presence of a cancerous condition within a tissue where less than all cells exhibit the complete pattern. Thus, for example, a set of selected genes, corresponding to any of the genes disclosed herein, may be found, using appropriate probes, either DNA or RNA, to be present in as little as 60% of cells derived from a sample of tumorous, or malignant, tissue while being absent from as much as 60% of cells derived from corresponding non-cancerous, or otherwise normal, tissue (and thus being present in as much as 40% of such normal tissue cells). In a

preferred embodiment, such gene pattern is found to be present in at least 50% of cells drawn from a cancerous tissue, such as the lung cancer disclosed herein. In an additional embodiment, such gene pattern is found to be present in at least 100% of cells drawn from a cancerous tissue and
5 absent from at least 100% of a corresponding normal, non-cancerous, tissue sample, although the latter embodiment may represent a rare occurrence.

In another aspect the present invention relates to a process for determining the cancerous status of a test cell, comprising determining
10 expression in said test cell of a gene as disclosed herein and then comparing said expression to expression of said at least one gene in at least one cell known to be non-cancerous whereby a difference in said expression indicates that said cell is cancerous.

15 In one embodiment, said change in expression is a change in copy number, including either an increase or decrease in copy number. In accordance with the present invention, said change in gene copy number may be determined by determining a change in expression of messenger RNA encoded by said gene.

20

Changes in gene copy number may be determined by determining a change in expression of messenger RNA encoded by a particular gene, especially that of Such change in gene copy number may be determined by determining a change in expression of messenger RNA encoded by a
25 particular gene, especially that of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2. Also in accordance with the present invention, said gene may be a cancer initiating gene, a cancer facilitating gene, or a cancer suppressing gene. In carrying out the
30 methods of the present invention, a cancer facilitating gene is a gene that, while not directly initiating or suppressing tumor formation or growth, said gene acts, such as through the actions of its expression product, to direct,

enhance, or otherwise facilitate the progress of the cancerous condition, including where such gene acts against genes, or gene expression products, that would otherwise have the effect of decreasing tumor formation and/or growth.

5

Although the presence or absence of expression of a gene corresponding to one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2, may be indicative of a cancerous status for a given cell, the mere presence or absence of such a gene may not alone be sufficient to achieve a malignant condition and thus the level of expression of such gene pattern may also be a significant factor in determining the attainment of a cancerous state. Thus, while a pattern of genes may be present in both cancerous and non-cancerous cells, the level of expression, as determined by any of the methods disclosed herein, all of which are well known in the art, may differ between the cancerous versus the non-cancerous cells. Thus, it becomes essential to also determine the level of expression of a gene such as that disclosed herein, including substantially similar genes, as a separate means of diagnosing the presence of a cancerous status for a given cell, groups of cells, or tissues, either in culture or *in situ*.

The level of expression of the polypeptides disclosed herein is also a measure of gene expression, such as polypeptides having sequence identical, or similar to any polypeptide encoded by any of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2.

Thus, the present invention further relates to a method for identifying an agent that modulates the activity of a cancer-target polypeptide comprising:

(a) contacting a test compound with a cell expressing a polypeptide encoded by a polynucleotide corresponding to a gene having the properties of a, b and c disclosed above for identifying a cancer-target gene and under conditions promoting the expression of said polypeptide; and

5 (b) determining a difference in expression of said polypeptide relative to when said test compound is not present wherein said difference indicates cancer-target polypeptide modulating activity,

thereby identifying a cancer-target polypeptide modulating agent.

10 The present invention further relates to a method for identifying an agent that modulates the activity of a cancer-target polypeptide comprising:

(a) contacting a test compound with a polypeptide encoded by a polynucleotide corresponding to a gene having the properties of a, b and c of claim 1 and under conditions promoting the activity of said polypeptide; and

15 (b) determining a difference in activity of said polypeptide relative to when said test compound is not present wherein said difference indicates cancer-target polypeptide modulating activity,

thereby identifying a cancer-target polypeptide modulating agent.

20 In any of these methods, a preferred embodiment utilizes a gene selected from KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*.

25 In accordance with the foregoing, the present invention further relates to a process for determining the cancerous status of a cell to be tested, comprising determining the level of expression in said cell of at least one gene of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*,
30 *PRO2000*, *RFC2* and *RIPK2*, including genes substantially identical to said sequences, or characteristic fragments thereof, or the complements of any of the foregoing and then comparing said expression to that of a cell known to

be non-cancerous whereby the difference in said expression indicates that said cell to be tested is cancerous.

In accordance with the invention, although gene expression for a gene useful in the methods of the invention is preferably determined by use of a probe that is a fragment of such nucleotide sequence, it is to be understood that the probe may be formed from a different portion of the gene. Expression of the gene may be determined by use of a nucleotide probe that hybridizes to messenger RNA (mRNA) transcribed from a portion of the gene.

It should be noted that there are a variety of different contexts in which genes have been evaluated as being involved in the cancerous process. Thus, some genes may be oncogenes and encode proteins that are directly involved in the cancerous process and thereby promote the occurrence of cancer in an animal. In addition, other genes may serve to suppress the cancerous state in a given cell or cell type and thereby work against a cancerous condition forming in an animal. Other genes may simply be involved either directly or indirectly in the cancerous process or condition and may serve in an ancillary capacity with respect to the cancerous state. All such types of genes are deemed with those to be determined in accordance with the invention as disclosed herein. Thus, the gene determined by said process of the invention may be an oncogene, or the gene determined by said process may be a cancer facilitating gene, the latter including a gene that directly or indirectly affects the cancerous process, either in the promotion of a cancerous condition or in facilitating the progress of cancerous growth or otherwise modulating the growth of cancer cells, either *in vivo* or *ex vivo*. In addition, the gene determined by said process may be a cancer suppressor gene, which gene works either directly or indirectly to suppress the initiation or progress of a cancerous condition. Such genes may work indirectly where their expression alters the activity of some other gene or gene expression product that is itself directly involved in initiating or facilitating the progress of a cancerous condition. For example, a gene that encodes a polypeptide,

either wild or mutant in type, which polypeptide acts to suppress of tumor suppressor gene, or its expression product, will thereby act indirectly to promote tumor growth.

5 In accordance with the foregoing, the methods of the present invention includes cancer modulating agents that are themselves either polypeptides, or small chemical entities, that affect the cancerous process, including initiation, suppression or facilitation of tumor growth, either *in vivo* or *ex vivo*. Said cancer modulating agent may have the effect of increasing gene expression
10 or said cancer modulating agent may have the effect of decreasing gene expression as such terms have been described herein.

 Thus, the present invention relates to a method for treating cancer comprising contacting a cancerous cell with an effective amount of an agent
15 that can reduce the activity of a cancer-target gene (i.e., a gene having the properties of a, b and c disclosed herein for identifying a cancer-target gene). In a preferred embodiment, said gene is one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*.

20

 Such embodiments include use of any of the agents having activity in one or more of the screening methods disclosed herein, most preferably wherein the agent was first identified as having such activity using one or more of said methods. In a preferred embodiment, the cancerous cell is
25 contacted *in vivo*.

25

 In an additional preferred embodiment, the agent has affinity for an expression product of said gene, such as where the agent is an antibody, preferably one disclosed according to the present invention.

30

 The proteins encoded by the genes disclosed herein due to their expression, or elevated expression, in cancer cells, represent highly useful

therapeutic targets for "targeted therapies" utilizing such affinity structures as, for example, antibodies coupled to some cytotoxic agent. In such methodology, it is advantageous that nothing need be known about the endogenous ligands or binding partners for such cell surface molecules.

- 5 Rather, an antibody or equivalent molecule that can specifically recognize the cell surface molecule (which could include an artificial peptide, a surrogate ligand, and the like) that is coupled to some agent that can induce cell death or a block in cell cycling offers therapeutic promise against these proteins. Thus, such approaches include the use of so-called suicide "bullets" against
10 intracellular proteins

- With the advent of methods of molecular biology and recombinant technology, it is now possible to produce antibody molecules by recombinant means and thereby generate gene sequences that code for
15 specific amino acid sequences found in the polypeptide structure of the antibodies. Such antibodies can be produced by either cloning the gene sequences encoding the polypeptide chains of said antibodies or by direct synthesis of said polypeptide chains, with *in vitro* assembly of the synthesized chains to form active tetrameric (H_2L_2) structures with affinity
20 for specific epitopes and antigenic determinants. This has permitted the ready production of antibodies having sequences characteristic of neutralizing antibodies from different species and sources.

- Regardless of the source of the antibodies, or how they are
25 recombinantly constructed, or how they are synthesized, *in vitro* or *in vivo*, using transgenic animals, such as cows, goats and sheep, using large cell cultures of laboratory or commercial size, in bioreactors or by direct chemical synthesis employing no living organisms at any stage of the process, all antibodies have a similar overall 3 dimensional structure.
30 This structure is often given as H_2L_2 and refers to the fact that antibodies commonly comprise 2 light (L) amino acid chains and 2 heavy (H) amino

acid chains. Both chains have regions capable of interacting with a structurally complementary antigenic target. The regions interacting with the target are referred to as "variable" or "V" regions and are characterized by differences in amino acid sequence from antibodies of
5 different antigenic specificity.

The variable regions of either H or L chains contains the amino acid sequences capable of specifically binding to antigenic targets. Within these sequences are smaller sequences dubbed "hypervariable" because
10 of their extreme variability between antibodies of differing specificity. Such hypervariable regions are also referred to as "complementarity determining regions" or "CDR" regions. These CDR regions account for the basic specificity of the antibody for a particular antigenic determinant structure.

15

The CDRs represent non-contiguous stretches of amino acids within the variable regions but, regardless of species, the positional locations of these critical amino acid sequences within the variable heavy and light chain regions have been found to have similar locations within
20 the amino acid sequences of the variable chains. The variable heavy and light chains of all antibodies each have 3 CDR regions, each non-contiguous with the others (termed L1, L2, L3, H1, H2, H3) for the respective light (L) and heavy (H) chains. The accepted CDR regions have been described by Kabat et al, *J. Biol. Chem.* 252:6609-6616 (1977).
25 The numbering scheme is shown in the figures, where the CDRs are underlined and the numbers follow the Kabat scheme.

In all mammalian species, antibody polypeptides contain constant (i.e., highly conserved) and variable regions, and, within the latter, there
30 are the CDRs and the so-called "framework regions" made up of amino

acid sequences within the variable region of the heavy or light chain but outside the CDRs.

The antibodies disclosed according to the invention may also be wholly synthetic, wherein the polypeptide chains of the antibodies are synthesized and, possibly, optimized for binding to the polypeptides disclosed herein as being receptors. Such antibodies may be chimeric or humanized antibodies and may be fully tetrameric in structure, or may be dimeric and comprise only a single heavy and a single light chain. Such antibodies may also include fragments, such as Fab and F(ab₂)' fragments, capable of reacting with and binding to any of the polypeptides disclosed herein as being receptors.

In one aspect, the present invention relates to immunoglobulins, or antibodies, as described herein, that react with, especially where they are specific for, the polypeptides encoded by a gene identified by the methods of the invention, preferably one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*. Such antibodies may commonly be in the form of a composition, especially a pharmaceutical composition.

Thus, the present invention contemplates an antibody that binds to a polypeptide encoded by a cancer-target gene (i.e., a gene having the properties of a, b and c disclosed above for identifying a cancer-target gene). In a preferred embodiment, the polypeptide or protein is encoded by one or more of genes KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*. In specific embodiments, this may include a naturally occurring antibody, such as polyclonal antibodies, or more preferably a monoclonal antibody, or a recombinant antibody or a partly or wholly synthetic antibody. In additional preferred embodiments, the antibody further comprises a cytotoxic agent, such as an apoptotic agent.

The pharmaceutical compositions useful herein also contain a pharmaceutically acceptable carrier, including any suitable diluent or excipient, which includes any pharmaceutical agent that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Pharmaceutically acceptable carriers include, but are not limited to, liquids such as water, saline, glycerol and ethanol, and the like, including carriers useful in forming sprays for nasal and other respiratory tract delivery or for delivery to the ophthalmic system. A thorough discussion of pharmaceutically acceptable carriers, diluents, and other excipients is presented in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. current edition).

The process of the present invention includes embodiments of the above-recited processes wherein the cancer cell is contacted *in vivo* as well as *ex vivo*, preferably wherein said agent comprises a portion, or is part of an overall molecular structure, having affinity for said expression product. In one such embodiment, said portion having affinity for said expression product is an antibody, especially where said expression product is a polypeptide or oligopeptide or comprises an oligopeptide portion, or comprises a polypeptide.

Such an agent can therefore be a single molecular structure, comprising both affinity portion and anti-cancer activity portions, wherein said portions are derived from separate molecules, or molecular structures, possessing such activity when separated and wherein such agent has been formed by combining said portions into one larger molecular structure, such as where said portions are combined into the form of an adduct. Said anti-cancer and affinity portions may be joined covalently, such as in the form of a single polypeptide, or polypeptide-like, structure or may be joined non-covalently, such as by hydrophobic or electrostatic interactions, such structures having been formed by means well known in the chemical arts.

Alternatively, the anti-cancer and affinity portions may be formed from separate domains of a single molecule that exhibits, as part of the same chemical structure, more than one activity wherein one of the activities is against cancer cells, or tumor formation or growth, and the other activity is
5 affinity for an expression product produced by expression of genes related to the cancerous process or condition.

In one embodiment of the present invention, a chemical agent, such as a protein or other polypeptide, is joined to an agent, such as an antibody,
10 having affinity for an expression product of a cancerous cell, such as a polypeptide or protein encoded by a gene related to the cancerous process, especially a gene as disclosed herein according to the present invention. Thus, where the presence of said expression product is essential to tumor initiation and/or growth, binding of said agent to said expression product will
15 have the effect of negating said tumor promoting activity. In one such embodiment, said agent is an apoptosis-inducing agent that induces cell suicide, thereby killing the cancer cell and halting tumor growth..

Other genes within the cancer cell that are regulated in a manner
20 similar to that of the genes disclosed herein and thus change their expression in a coordinated way in response to chemical compounds represent genes that are located within a common metabolic, signaling, physiological, or functional pathway so that by analyzing and identifying such commonly regulated groups of genes (groups that include the gene, or similar
25 sequences, disclosed according to the invention, one can (a) assign known genes and novel genes to specific pathways and (b) identify specific functions and functional roles for novel genes that are grouped into pathways with genes for which their functions are already characterized or described. For example, one might identify a group of 10 genes, at least one of which is the
30 gene as disclosed herein, that change expression in a coordinated fashion and for which the function of one, such as the polypeptide encoded by a gene disclosed herein, is known then the other genes are thereby implicated in a

similar function or pathway and may thus play a role in the cancer-initiating or cancer-facilitating process. In the same way, if a gene were found in normal cells but not in cancer cells, or happens to be expressed at a higher level in normal as opposed to cancer cells, then a similar conclusion may be drawn as to its involvement in cancer, or other diseases. Therefore, the processes disclosed according to the present invention at once provide a novel means of assigning function to genes, i.e. a novel method of functional genomics, and a means for identifying chemical compounds that have potential therapeutic effects on specific cellular pathways. Such chemical compounds may have therapeutic relevance to a variety of diseases outside of cancer as well, in cases where such diseases are known or are demonstrated to involve the specific cellular pathway that is affected.

The polypeptides contemplated by the invention, preferably those encoded by one or more of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*, also find use as vaccines in that, where the polypeptide represents a surface protein present on a cancer cell, such polypeptide may be administered to an animal, especially a human being, for purposes of activating cytotoxic T lymphocytes (CTLs) that will be specific for, and act to lyse, cancer cells in said animal. Where used as vaccines, such polypeptides are present in the form of a pharmaceutical composition. The present invention may also employ polypeptides that have the same, or similar, immunogenic character as the polypeptides encoded by one or more of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2*, and thereby elicit the same, or similar, immunogenic response after administration to an animal, such as an animal at risk of developing cancer, or afflicted therewith. Thus, the polypeptides disclosed according to the invention will commonly find use as immunogenic compositions.

The present invention also relates to a process that comprises a method for producing a product, such as by generating test data to facilitate identification of such product, comprising identifying an agent according to one of the disclosed processes for identifying such an agent (i.e., the therapeutic agents identified according to the assay procedures disclosed herein) wherein said product is the data collected with respect to said agent as a result of said identification process, or assay, and wherein said data is sufficient to convey the chemical character and/or structure and/or properties of said agent. For example, the present invention specifically contemplates a situation whereby a user of an assay of the invention may use the assay to screen for compounds having the desired enzyme modulating activity and, having identified the compound, then conveys that information (i.e., information as to structure, dosage, etc) to another user who then utilizes the information to reproduce the agent and administer it for therapeutic or research purposes according to the invention. For example, the user of the assay (user 1) may screen a number of test compounds without knowing the structure or identity of the compounds (such as where a number of code numbers are used the first user is simply given samples labeled with said code numbers) and, after performing the screening process, using one or more assay processes of the present invention, then imparts to a second user (user 2), verbally or in writing or some equivalent fashion, sufficient information to identify the compounds having a particular modulating activity (for example, the code number with the corresponding results). This transmission of information from user 1 to user 2 is specifically contemplated by the present invention.

The genes useful in the methods of the invention disclosed herein are genes corresponding to one of KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, ARHC, CDC6, CDK7, CDKN3, CRK7, DUSP16, FIGNL1, GUK1, ITPR2, KCNK1, KCNK5, PRO2000, RFC2 and RIPK2 and represent genes

that may be over-expressed in malignant cancer. In addition, in any given sample, not all cancer cells may express this gene a substantial expression thereof in a substantial number of such cells is sufficient to warrant a determination of a cancerous, or potentially cancerous, condition.

5

Thus, the genes disclosed according to the present invention are expressed in cancer compared to normal tissue samples or may be expressed at a higher level in cancer as compared to normal tissues. Further, such polynucleotide, or gene, sequence expression in normal tissues may correlate with individuals having a family history of cancer.

10

Such genes may play a direct role in cancer progression, such as in cancer initiation or cancer cell proliferation/survival. For example, one or more genes encoding the same polypeptide as one or more of the sequences disclosed herein represent novel individual gene targets for screening and discovery of small molecules that inhibit enzyme or other cellular functions, e.g. kinase inhibitors. Such molecules represent valuable therapeutics for cancer. In addition, small molecules or agents, such as small organic molecules, that down-regulate the expression of these genes in cancer would represent valuable anti-cancer therapeutics. Expression of the gene in normal tissues may indicate a predisposition towards development of lung cancer. The encoded polypeptide might represent a potentially useful cell surface target for therapeutic molecules such as cytolytic antibodies, or antibodies attached to cytotoxic, or cytolytic, agents.

20
25

Expression of a gene corresponding to a polynucleotide disclosed herein, when in normal tissues, may indicate a predisposition towards development of colon cancer. The encoded polypeptide might then present a potentially useful cell surface target for therapeutic molecules such as cytolytic antibodies, or antibodies attached to cytotoxic, or cytolytic, agents.

30

The present invention specifically contemplates use of antibodies against polypeptides encoded by the genes disclosed herein, whereby said antibodies are conjugated to one or more cytotoxic agents so that the antibodies serve to target the conjugated immunotoxins to a region of cancerous activity, such as a solid tumor. For many known cytotoxic agents, lack of selectivity has presented a drawback to their use as therapeutic agents in the treatment of malignancies. For example, the class of two-chain toxins, consisting of a binding subunit (or B-chain) linked to a toxic subunit (A-chain) are extremely cytotoxic. Thus, such agents as ricin, a protein isolated from castor beans, kills cells at very low concentrations (even less than 10^{-11} M) by inactivating ribosomes in said cells (see, for example, Lord et al., Ricin: structure, mode of action, and some current applications. *Faseb J*, 8: 201-208 (1994), and Blättler et al., Realizing the full potential of immunotoxins. *Cancer Cells*, 1: 50-55 (1989)). While isolated A-chains of protein toxins that functionally resemble ricin A-chain are only weakly cytotoxic for intact cells (in the concentration range of 10^{-7} to 10^{-6} M), they are very potent cytotoxic agents inside the cells. Thus, a single molecule of the A-subunit of diphtheria toxin can kill a cell once inside (see: Yamaizumi et al., One molecule of diphtheria toxin fragment A introduced into a cell can kill the cell. *Cell*, 15: 245-250, 1978).

The present invention solves this selectivity problem by using antibodies specific for antigens present on cancer cells to target the cytotoxins to said cells. In addition, use of antibodies decreases toxicity because the antibodies are non-toxic until they reach the tumor and, because the cytotoxin is bound to the antibody, it is presented with less opportunity to cause damage to non-targeted tissues.

In addition, use of such antibodies alone can provide therapeutic effects on the tumor through the antibody-dependent cellular cytotoxic response (ADCC) and complement-mediated cell lysis mechanisms.

A number of recombinant immunotoxins (for example, consisting of Fv regions of cancer specific antibodies fused to truncated bacterial toxins) are well known (see, for example, Smyth et al., Specific targeting of chlorambucil to tumors with the use of monoclonal antibodies, *J. Natl. Cancer Inst.*, 5 76(3):503-510 (1986); Cho et al., Single-chain Fv/folate conjugates mediate efficient lysis of folate-receptor-positive tumor cells, *Bioconjug. Chem.*, 8(3):338-346 (1997)). As noted in the literature, these may contain, for example, a truncated version of *Pseudomonas* exotoxin as a toxic moiety but the toxin is modified in such a manner that by itself it does not bind to normal 10 human cells, but it retains all other functions of cytotoxicity. Here, recombinant antibody fragments target the modified toxin to cancer cells which are killed, such as by direct inhibition of protein synthesis, or by concomitant induction of apoptosis. Cells that are not recognized by the antibody fragment, because they do not carry the cancer antigen, are not affected. Good activity and 15 specificity has been observed for many recombinant immunotoxins in *in vitro* assays using cultured cancer cells as well as in animal tumor models. Ongoing clinical trials provide examples where the promising pre-clinical data correlate with successful results in experimental cancer therapy. (see, for example, Brinkmann U., Recombinant antibody fragments and immunotoxin 20 fusions for cancer therapy, *In Vivo* (2000) 14:21-27).

While the safety of employing immunoconjugates in humans has been established, *in vivo* therapeutic results have been less impressive. Because clinical use of mouse MAbs in humans is limited by the development of a 25 foreign anti-globulin immune response by the human host, genetically engineered chimeric human-mouse MAbs have been developed by replacing the mouse Fc region with the human constant region. In other cases, the mouse antibodies have been "humanized" by replacing the framework regions of variable domains of rodent antibodies by their human equivalents. Such 30 humanized and engineered antibodies can even be structurally arranged to have specificities and effector functions determined by design and which characteristics do not appear in nature. The development of bispecific

antibodies, having different binding ends so that more than one antigenic site can be bound, have proven useful in targeting cancer cells. Thus, such antibody specificity has been improved by chemical coupling to various agents such as bacterial or plant toxins, radionuclides or cytotoxic drugs and other agents. (see, for example, Bodey, B. et al). Genetically engineered monoclonal antibodies for direct anti-neoplastic treatment and cancer cell specific delivery of chemotherapeutic agents. *Curr Pharm Des* (2000) Feb;6(3):261-76). See also, Garnett, M. C., Targeted drug conjugates: principles and progress. *Adv. Drug Deliv. Rev.* (2001 Dec 17) 53(2):171-216; Brinkmann et al., Recombinant immunotoxins for cancer therapy. *Expert Opin Biol Ther.* (2001) 1(4):693-702.

Among the cytotoxic agents specifically contemplated for use as immunoconjugates according to the present invention are Calicheamicin, a highly toxic enediyne antibiotic isolated from *Micromonospora echinospora ssp. Calichensis*, and which binds to the minor groove of DNA to induce double strand breaks and cell death (see: Lee et al., Calicheamicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calicheamicin g₁. *J Am Chem Soc*, 109: 3464-3466 (1987); Zein et al., Calicheamicin gamma 1l: an antitumor antibiotic that cleaves double-stranded DNA site specifically, *Science*, 240: 1198-1201 (1988)). Useful derivatives of the calicheamicins include mylotarg and 138H11-Cam θ . Mylotarg is an immunoconjugate of a humanized anti-CD33 antibody (CD33 being found in leukemic cells of most patients with acute myeloid leukemia) and N-acetyl gamma colicheamicin dimethyl hydrazide, the latter of which is readily coupled to an antibody of the present invention (in place of the anti-CD33 but which can also be humanized by substitution of human framework regions into the antibody during production as described elsewhere herein) to form an immunoconjugate of the invention. (see: Hamann et al. Gemtuzumab Ozogamicin, A Potent and Selective Anti-CD33 Antibody-Calicheamicin Conjugate for Treatment of Acute Myeloid Leukemia, *Bioconjug. Chem.* 13, 47-58 (2002)) For use with 138H11-Cam θ , 138H11 is an anti- γ -glutamyl

transferase antibody coupled to theta calicheamicin through a disulfide linkage and found useful *in vitro* against cultured renal cell carcinoma cells. (see: Knoll et al., Targeted therapy of experimental renal cell carcinoma with a novel conjugate of monoclonal antibody 138H11 and calicheamicin θ_1^I ,
5 *Cancer Res*, **60**: 6089-6094 (2000)) The same linkage may be utilized to link this cytotoxic agent to an antibody of the present invention, thereby forming a targeting structure for colon cancer cells.

Also useful in forming the immunoconjugates of the invention is DC1, a
10 disulfide-containing analog of adozelesin, that kills cells by binding to the minor groove of DNA, followed by alkylation of adenine bases. Adozelesin is a structural analog of CC-1065, an anti-tumor antibiotic isolated from microbial fermentation of *Streptomyces zelensis*, and is about 1,000 fold more toxic to
15 cultured cell lines than other DNA interacting agents, such as cis-platin and doxorubicin. This agent is readily linked to antibodies through the disulfide bond of adozelesin. (see: Chari et al., Enhancement of the selectivity and antitumor efficacy of a CC-1065 analogue through immunoconjugate formation, *Cancer Res*, **55**: 4079-4084 (1995)).

20 Maytansine, a highly cytotoxic microtubular inhibitor isolated from the shrub *Maytenus serrata* found to have little value in human clinical trials, is much more effective in its derivatized form, denoted DM1, containing a disulfide bond to facilitate linkage to antibodies, is up to 10-fold more cytotoxic (see: Chari et al., Immunoconjugates containing novel maytansinoids:
25 promising anticancer drugs, *Cancer Res*, **52**: 127-131 (1992)). These same *in vitro* studies showed that up to four DM1 molecules could be linked to a single immunoglobulin without destroying the binding affinity. Such conjugates have been used against breast cancer antigens, such as the *neu*/HER2/*erbB-2* antigen. (see: Goldmacher et al., Immunogen, Inc., (2002) *in press*); also see
30 Liu, C. et al., Eradication of large colon tumor xenografts by targeted delivery of maytansinoids, *Proc. Natl. Acad. Sci. USA*, **93**, 8618-8623 (1996)). For example, Liu et al. (1996) describes formation of an immunoconjugate of the

maytansinoid cytotoxin DM1 and C242 antibody, a murine IgG1 immunoglobulin, available from Pharmacia and which has affinity for a mucin-like glycoprotein variably expressed by human colorectal cancers. The latter immunoconjugate was prepared according to Chari et al., *Cancer Res.*,
5 52:127-131 (1992) and was found to be highly cytotoxic against cultured colon cancer cells as well as showing anti-tumor effects *in vivo* in mice bearing subcutaneous COLO 205 human colon tumor xenografts using doses well below the maximum tolerated dose.

10 In accordance with the foregoing, a preferred embodiment of the present invention includes where the cytotoxic agent is a calicheamicin, a maytansinoid, an adozelesin, DC1, a cytotoxic protein, a taxol, a taxotere, or a taxoid. In especially preferred embodiments, the calicheamicin is calicheamicin γ_1^I , N-acetyl gamma calicheamicin dimethyl hydrazide or
15 calicheamicin θ_1^I , the maytansinoid is DM1, the cytotoxic protein is ricin, abrin, gelonin, pseudomonas exotoxin or diphtheria toxin, the taxol is paclitaxel, and the taxotere is docetaxel.

In addition, there are a variety of protein toxins (cytotoxic proteins),
20 which include a number of different classes, such as those that inhibit protein synthesis: ribosome-inactivating proteins of plant origin, such as ricin, abrin, gelonin, and a number of others, and bacterial toxins such as pseudomonas exotoxin and diphtheria toxin.

25 Another useful class is the one including taxol, taxotere, and taxoids. Specific examples include paclitaxel (taxol), its analog docetaxel (taxotere), and derivatives thereof. The first two are clinical drugs used in treating a number of tumors while the taxoids act to induce cell death by inhibiting the de-polymerization of tubulin. Such agents are readily linked to antibodies
30 through disulfide bonds without disadvantageous effects on binding specificity.

In one instance, a truncated *Pseudomonas* exotoxin was fused to an anti-CD22 variable fragment and used successfully to treat patients with chemotherapy-resistant hairy-cell leukemia. (see: Kreitman et al., Efficacy of the anti-CD22 recombinant immunotoxin BL22 in chemotherapy-resistant hairy-cell leukemia, *N Engl J Med*, **345**: 241-247 (2001)) Conversely, the cancer-linked peptides of the present invention offer the opportunity to prepare antibodies, recombinant or otherwise, against the appropriate antigens to target solid tumors, preferably those of malignancies of colon tissue, using the same or similar cytotoxic conjugates. Thus, many of the previously used immunoconjugates have been formed using antibodies against general antigenic sites linked to cancers whereas the antibodies formed using the peptides disclosed herein are more specific and target the antibody-cytotoxic agent to a particular tissue or organ, thus further reducing toxicity and other undesirable side effects.

15

In addition, the immunoconjugates formed using the antibodies prepared against the cancer-linked antigens disclosed herein can be formed by any type of chemical coupling. Thus, the cytotoxic agent of choice, along with the immunoglobulin, can be coupled by any type of chemical linkage, covalent or non-covalent, including electrostatic linkage, to form the immunoconjugates of the present invention.

20

When used as immunoconjugates, the antitumor agents of the present invention represent a class of pro-drugs that are relatively non-toxic when first administered to an animal (due mostly to the stability of the immunoconjugate), such as a human patient, but which are targeted by the conjugated immunoglobulin to a cancer cell where they then exhibit good toxicity. The tumor-related, associated, or linked, antigens, preferably those presented herein, serve as targets for the antibodies (monoclonal, recombinant, and the like) specific for said antigens. The end result is the release of active cytotoxic agent inside the cell after binding of the immunoglobulin portion of the immunoconjugate.

25

30

The cited references describe a number of useful procedures for the chemical linkage of cytotoxic agents to immunoglobulins and the disclosures of all such references cited herein are hereby incorporated by reference in their entirety. For other reviews see Ghetie et al., Immunotoxins in the therapy of cancer: from bench to clinic, *Pharmacol Ther*, **63**: 209-234 (1994), Pietersz et al. The use of monoclonal antibody immunoconjugates in cancer therapy, *Adv Exp Med Biol*, **353**:169-179 (1994), and Pietersz, G. A. The linkage of cytotoxic drugs to monoclonal antibodies for the treatment of cancer, *Bioconjug Chem*, **1**:89-95 (1990).

Thus, the present invention provides highly useful cancer-associated antigens for generation of antibodies for linkage to a number of different cytotoxic agents which are already known to have some *in vitro* toxicity and possess chemical groups available for linkage to antibodies.

It should be cautioned that, in carrying out the procedures of the present invention as disclosed herein, any reference to particular buffers, media, reagents, cells, culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

30

EXAMPLE

SW480 cells are grown to a density of 10^5 cells/cm² in Leibovitz's L-15 medium supplemented with 2 mM L-glutamine (90%) and 10% fetal bovine serum. The cells are collected after treatment with 0.25% trypsin, 0.02% EDTA at 37°C for 2 to 5 minutes. The trypsinized cells are then diluted with 30 ml growth medium and plated at a density of 50,000 cells per well in a 96 well plate (200 μ l/well). The following day, cells are treated with either compound buffer alone, or compound buffer containing a chemical agent to be tested, for 24 hours. The media is then removed, the cells lysed and the RNA recovered using the RNeasy reagents and protocol obtained from Qiagen. RNA is quantitated and 10 ng of sample in 1 μ l are added to 24 μ l of Taqman reaction mix containing 1X PCR buffer, RNAsin, reverse transcriptase, nucleoside triphosphates, amplitaq gold, tween 20, glycerol, bovine serum albumin (BSA) and specific PCR primers and probes for a reference gene (18S RNA) and a test gene (Gene X). Reverse transcription is then carried out at 48°C for 30 minutes. The sample is then applied to a Perlin Elmer 7700 sequence detector and heat denatured for 10 minutes at 95°C. Amplification is performed through 40 cycles using 15 seconds annealing at 60°C followed by a 60 second extension at 72°C and 30 second denaturation at 95°C. Data files are then captured and the data analyzed with the appropriate baseline windows and thresholds.

The quantitative difference between the target and reference gene is then calculated and a relative expression value determined for all of the samples used. This procedure is then repeated for other genes functionally related to the gene as disclosed herein and the level of function, or expression, noted. The relative expression ratios for each pair of genes is determined (i.e., a ratio of expression is determined for each target gene versus each of the other genes for which expression is measured, where each gene's absolute expression is determined relative to the reference gene for each compound, or chemical agent, to be screened). The samples are then

scored and ranked according to the degree of alteration of the expression profile in the treated samples relative to the control. The overall expression of the particular gene relative to the controls, as modulated by one chemical agent relative to another, is also ascertained. Chemical agents having the most effect on a given gene, or set of genes, are considered the most anti-neoplastic.

Table 6 below contains a listing of the genes (numbered 1 to 20) along with their Gencarta names (or accessions). Each gene is represented as a consensus sequence followed by predicted mRNA transcripts and then predicted polypeptides. All of the sequences, with additional information, are presented in Figure 1.

The present invention also relates to an isolated cancer target gene wherein said gene is a gene identified in Table 6. Thus, the present invention encompasses isolated genes identified herein as KIAA1274, NEK6, PAK2, PAK4, STK38L, ACP1, *ARHC*, *CDC6*, *CDK7*, *CDKN3*, *CRK7*, *DUSP16*, *FIGNL1*, *GUK1*, *ITPR2*, *KCNK1*, *KCNK5*, *PRO2000*, *RFC2* and *RIPK2* and uses of these genes, whether isolated or not, in any of the methods of the invention.

Table 6. Sequence Identification Numbers for genes, transcripts and polypeptides*

Gene	Accession No.	Consensus	Transcripts	Polypeptides
1	AA383349	1	2-10	11-16
2	AA553584	17	18-22	23-27
3	D61791	28	29-36	37-40
4	F02366	41	42-56	57-68
5	H61320	69	70-75	76-78
6	HUMAAPA	79	80-104	105-115
7	HUMPTPB	116	117-125	126-131
8	R03897	132	133-150	151-158
9	R14324	159	160-165	166-170
10	R25184	171	172-176	177-179
11	T08090	180	181-235	236-254
12	T11445	255	256-282	283-294
13	T23935	295	296-310	311-315
14	T60764	316	317-329	330-334
15	T62520	335	336-365	366-378
16	T83032	379	380-389	390-392
17	Z26993	393	394-415	416-421
18	Z38709	422	423-439	440-449
19	Z39663	450	451-461	462-470
20	Z44462	471	472-490	491-501

5 *Accession numbers in Table 6 are for the Gencarta database.